PRML Assignment-1

IIT Jammu

January 31, 2025

Submission Guidelines

- **Submission Deadline:** February 7, 2025, 11:59 PM.
- **File Format:** All submissions should be in PDF format.
- **Code Submission:** If any programming or plots are required, submit a zip file containing:
  - A Python script (`.py`).
  - Output graphs (if applicable) in PNG or PDF format.
- **Plagiarism Policy:**
  - Students are strictly prohibited from copying from each other.
  - All submissions will be checked for plagiarism using automated tools.
  - If plagiarism is detected, the submission will be marked as invalid.
- **Late Submission Policy:** Late submissions will not be accepted unless prior approval is granted.

Question 1: Bayes Classifier

Problem Statement

Consider a classification problem where we have two classes ω_1 and ω_2, with given prior probabilities and class-conditional probability distributions. You are provided with the following priors:

$$P(\omega_1) = 0.6, \quad P(\omega_2) = 0.4$$

The conditional probability densities for a feature x given a class are modeled as Gaussian distributions:

$$p(x|\omega_1) = \mathcal{N}(\mu_1, \sigma_1^2)$$

where $\mu_1 = 5$, $\sigma_1 = 1.5$.

$$p(x|\omega_2) = \mathcal{N}(\mu_2, \sigma_2^2)$$

where $\mu_2 = 8$, $\sigma_2 = 2.0$.

Tasks

1. Compute the Posterior Probabilities

Use Bayes theorem to compute $P(\omega_1|x)$ and $P(\omega_2|x)$ for values of x in the range $[0, 15]$.

Formula

$$P(\omega_j|x) = \frac{P(x|\omega_j)P(\omega_j)}{P(x)}$$

where:

$$P(x) = P(x|\omega_1)P(\omega_1) + P(x|\omega_2)P(\omega_2)$$

2. Plot the Graph

Plot the posterior probabilities $P(\omega_1|x)$ and $P(\omega_2|x)$ on the same graph. Mark the decision boundary where:

$$P(\omega_1|x) = P(\omega_2|x)$$

3. Decision Rule

Classify x as belonging to ω_1 if:

$$P(\omega_1|x) > P(\omega_2|x)$$

and vice versa. Shade the decision regions in the graph to clearly differentiate between the two classes.

Question 2: Minimum Error Rate Classification

Problem Statement

A factory uses an automated sorting system to classify objects into two categories, defective (ω_1) and non-defective (ω_2). The prior probabilities of these categories are:

$$P(\omega_1) = 0.3, \quad P(\omega_2) = 0.7$$

The probability density functions of the weight x for each class are given by:

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi(1.2)^2}} e^{-\frac{(x-4)^2}{2(1.2)^2}}$$

$$p(x|\omega_2) = \frac{1}{\sqrt{2\pi(1.5)^2}} e^{-\frac{(x-6)^2}{2(1.5)^2}}$$

Tasks

1. Compute the posterior probabilities $P(\omega_1|x)$ and $P(\omega_2|x)$ for $x \in [0, 10]$.
2. Determine the decision boundary where $P(\omega_1|x) = P(\omega_2|x)$.
3. Plot the class-conditional densities $p(x|\omega_1)$ and $p(x|\omega_2)$ on the same graph.
4. Plot the posterior probabilities $P(\omega_1|x)$ and $P(\omega_2|x)$ on another graph and mark the decision boundary.
5. Discuss the misclassification rate if the decision boundary is shifted slightly.

Question 3: Bayesian Decision Theory with Loss Function

Problem Statement

A bank uses a machine learning system to classify loan applicants into two risk categories:

$$\text{High Risk } (\omega_1) : P(\omega_1) = 0.4, \quad \text{Low Risk } (\omega_2) : P(\omega_2) = 0.6$$

The probability density functions for a feature **credit score x** given each class are modeled as:

$$p(x|\omega_1) = \frac{1}{\sqrt{2\pi(20)^2}} e^{-\frac{(x-550)^2}{2(20)^2}}$$

$$p(x|\omega_2) = \frac{1}{\sqrt{2\pi(25)^2}} e^{-\frac{(x-700)^2}{2(25)^2}}$$

A decision rule is used to minimize the expected loss. The loss function is:

Decision	**Actual Class ω_1**	**Actual Class ω_2**
Classify as High Risk (α_1)	0	2
Classify as Low Risk (α_2)	5	0

Tasks

1. Compute the posterior probabilities $P(\omega_1|x)$ and $P(\omega_2|x)$.
2. Compute the expected loss for both actions $R(\alpha_1|x)$ and $R(\alpha_2|x)$.
3. Find the decision boundary where the expected loss is equal for both actions.
4. Plot the expected loss functions $R(\alpha_1|x)$ and $R(\alpha_2|x)$.
5. Discuss how the decision boundary changes if the prior probability of high-risk applicants increases.

Question 4: Risk Minimization with Three Classes

Problem Statement

A self-driving car must classify objects into three categories:

- **Pedestrian** (ω_1)
- **Vehicle** (ω_2)
- **Empty Road** (ω_3)

The prior probabilities for each class are:

$$P(\omega_1) = 0.2, \quad P(\omega_2) = 0.5, \quad P(\omega_3) = 0.3$$

The feature vector x (distance from LiDAR sensor) follows normal distributions:

$$p(x|\omega_1) = \mathcal{N}(x; 2, 0.5)$$

$$p(x|\omega_2) = \mathcal{N}(x; 5, 1.0)$$

$$p(x|\omega_3) = \mathcal{N}(x; 10, 1.5)$$

A loss function is defined as follows:

Decision	**Pedestrian** (ω_1)	**Vehicle** (ω_2)	**Empty Road** (ω_3)
Predict Pedestrian (α_1)	0	5	10
Predict Vehicle (α_2)	3	0	7
Predict Empty Road (α_3)	8	6	0

Tasks

1. Compute the posterior probabilities $P(\omega_1|x)$, $P(\omega_2|x)$, and $P(\omega_3|x)$.
2. Compute the expected loss for each action.
3. Find the optimal decision boundary that minimizes the risk.
4. Plot the decision boundaries and label regions for different classes.
5. Discuss the effect of increasing the penalty for misclassifying a pedestrian.